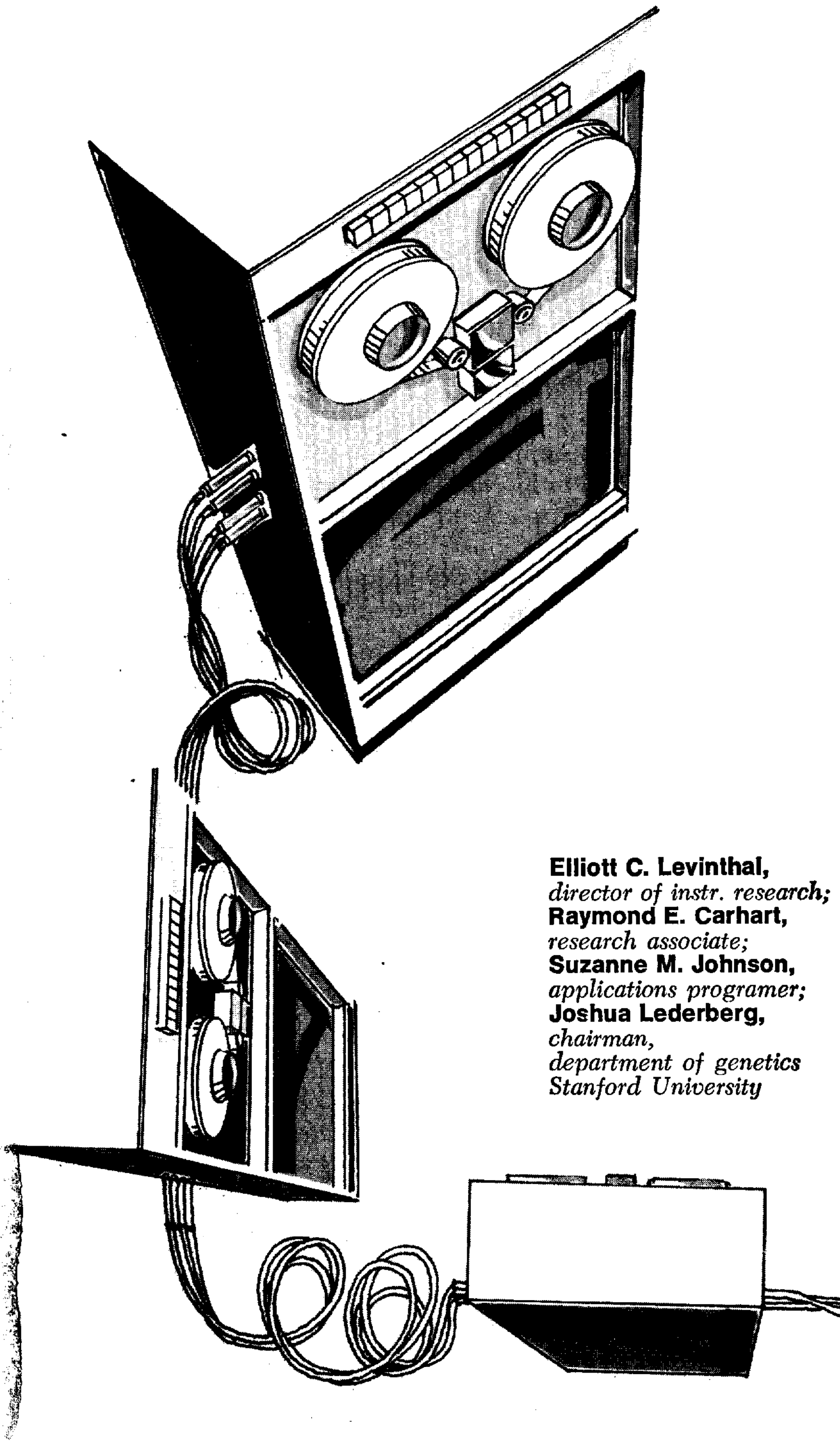
234

when computers 'talk' to computers

INDUSTRIAL RESEARCH—NOV 15, 1975
v. 17 (# 12): 35-42.



Elliott C. Levinthal,
director of instr. research;
Raymond E. Carhart,
research associate;
Suzanne M. Johnson,
applications programer;
Joshua Lederberg,
chairman,
department of genetics
Stanford University

THE EVER-GROWING POWER of computers and the complexity of the software this elicits, represent some of our most important contemporary scientific and technological frontiers. However, these same factors make it increasingly difficult to sustain a community of effective cooperation in the shared use of these intellectual efforts.

Publication in journal articles can do little but advertise the general capabilities of new design accomplishments in computer software. Scientists in different laboratories are frustrated in seeking to understand one another's efforts in the same spirit as applies to traditional fields of chemistry and biology. They are then driven to an "NIH" syndrome, *i. e.*, to close their minds to much that is "not invented here."

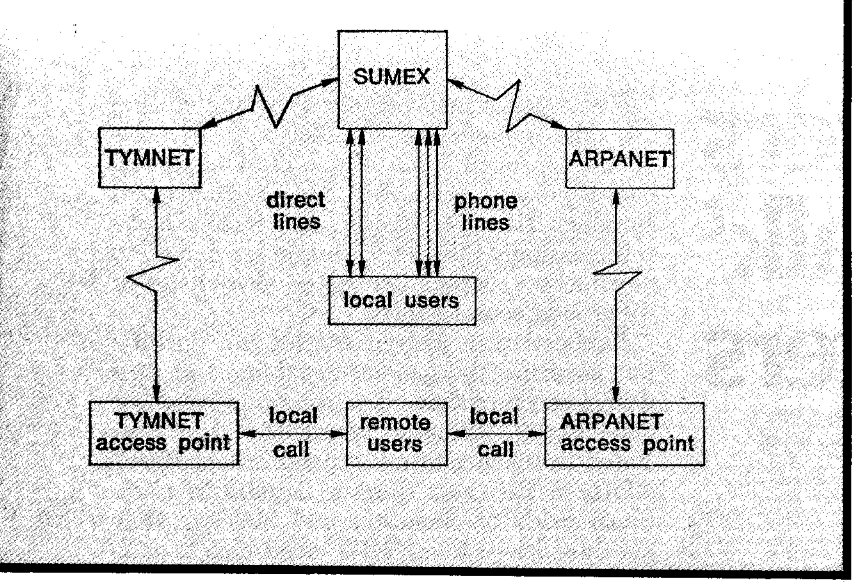
The promise of more-economical technologies for data communications arouses our hopes that geographical and operational barriers may be surmounted. Further efficiencies can be realized by the use of computer networking to create distributed sites at which functionally specialized capabilities are concentrated. The SUMEX-AIM (Stanford University Medical EXperimental computer—Artificial Intelligence in Medicine) project is an experiment in reducing this principle to practice, in the specific area of artificial intelligence research applied to health sciences.

The term "artificial intelligence" (AI) refers to research efforts aimed at studying and mechanizing information processing: tasks that generally have been considered to require human intelligence. The current emphasis in the field is to understand the underlying principles in efficient acquisition and utilization of material knowledge and respresentation of conceptual abstractions in reasoning, deductive, and problem-solving activities.

AI systems are characterized by complex computational processes that are primarily non-numeric, *e. g.*, graph-searching and symbolic pattern analysis. They involve procedures in which execution is controlled by different types and forms of knowledge about a given task domain, such as models, fragments of "advice," and systems of constraints or heuristic rules.

Unlike conventional algorithms commonly based on a well-tailored method for a given task, AI procedures typically use a multiplicity of methods in a highly conditional manner—depending on the specific data in the task and a variety of sources of relevant information. The tangible objective of this approach is the

a networking experiment in community research



Communication *with remote SUMEX-AIM projects is accomplished through one or more of the networks indicated above.*

production of computer programs. Using formal and informal knowledge together with mechanized hypothesis formation and problem-solving procedures, the approach will provide more general and effective consultative tools for the clinician and medical scientist.

These goals of SUMEX-AIM should be distinguished from the large number of other applications of computers to medicine and of computer networking. Computers are being applied to a wide range of problems of information management and communication involving patient histories and records, hospital and ambulatory care, and specific problems of clinical diagnostic and therapeutic instrumentation, such as EEG, radiation therapy and diagnosis, and laboratory analysis.

SUMEX-AIM is being directed to another arena—artificial intelligence:

☐ How do medical clinicians and basic scientists evolve solutions to the problems with which they are confronted?

☐ Can computers assist in these processes?

☐ Can they be used to extract "expertise" from experts and make it available to a larger community?

It also is important to distinguish SUMEX-AIM's networking objectives from many other computing networks. Most other networks stress their role as a utility. They often provide a set of software and hardware services to disparate users who do not share with each other. Alternatively the network can serve mainly as a communication system between separate groups of users whose members are geographically dispersed. SUMEX-AIM hopes to be a resource which the user contributes to as well as draws on. It seeks to serve as a medium of intellectual exchange.

The SUMEX-AIM computer facility (1) is a National Shared Computing Resource being developed and operated by Stanford University, in partnership with and with financial support from the Biotechnology Resources Branch of the Division of Research Resources, National Institutes of Health. It is national in scope in that a major portion of its computing capacity is being made available to authorized research groups throughout the country by means of communication networks.

In this article we discuss some problems and possible solutions related to networking and community-building (2).

The SUMEX-AIM facility

The SUMEX-AIM computer facility consists of a Digital Equipment Corp. model KI-10 central processor operating under the TENEX (14) time sharing monitor. It is located at Stanford University Medical Center, Stanford, Cal.

The system has 256K words (36 bit) of high speed memory; 1.6 x 10^6 words of swapping storage; 70 x 10^6 words of disk storage; two 9-track, 800 bpi industry-compatible tape units; one dual DEC-tape unit; a line printer; and communications network interfaces providing user terminal access via both TYMNET and ARPANET.

Software support has evolved, and will continue to evolve, based on user research goals and requirements. Major user languages currently include INTERLISP, SAIL (15), FORTRAN-10, BLISS-10, BASIC, and MACRO-10. Major software packages available include OMNIGRAPH, for graphics support of multiple terminal types, and MLAB, for mathematical modeling.

The SUMEX-AIM computer has limited operator support, as a cost-reducing measure; this does create some problems. Users who wish to run jobs requiring tapes must make arrangements to mount their own tapes. Likewise obtaining listings from the line printer may be delayed. Users generally rely on their own TTYs or displays for routine output.

Problems related to networking

During its first year of operation, the SUMEX-AIM facility has encountered a variety of problems arising from its network availability. In most cases, there has been no clear precedent for the handling of these situations. In fact, many problem-areas still reflect the influences of a yet-developing policy.

The hope is that this presentation and discussion of problems and their solutions may give foresight to others who contemplate networking or network use. The problems to be discussed can be loosely associated into three classes; those related to the management of the facility, those pertaining to research activities on the system, and those involving psychological barriers to network use.

"Gatekeeping." The most general problem faced by the organizers of the SUMEX-AIM facility is the question of "gatekeeping." In order to insure a high quality of pertinent research, some kind of refereeing system is needed to assess the value of proposed new projects. SUMEX-AIM is a national resource, half of its capacity being allocated by an independent national committee (the AIM Advisory Group). The remaining half is adminisstrated locally.

Choice of computer and operating system. A second management level problem is the choice of a computer and operating system which will optimize the usefulness of the facility for a majority of users, and which will encourage intercommunication between remote collaborators.

Because SUMEX-AIM is intended to be used primarily for applications of artificial intelligence, and because interactive LISP is a primary language in this type of work, the choice of TENEX as an operating system was strongly indicated TENEX incorporates multiple address spaces, thereby allowing multiple "fork" structure and paging, a design which is necessary to create the large-memory virtual machine required by existing verisions of INTERLISP (6).

The PDP-10 is a popular machine for interactive computing of all sorts in university research environments, and thus an added benefit of this choice was expected—the possibility of easily transferring to SUMEX programs developed at other sites. Many of these programs were written, not under TENEX, but under the 10/50 monitor supplied by the manufacturer. Because a large and useful program library was already available under the 10/50 monitor, one of the design criteria of TENEX was compatibility with such programs.

When a 10/50 program is run under TENEX, a special "compatibility package" of routines is invoked to translate 10/50 monitor calls into equivalent TENEX monitor calls. Considerable effort has been required to achieve a level of compatibility that still entails a moderate amount of hand tuning for complex programs.

TENEX provides many features that facilitate a comfortable network environment. The standard support programs included with this system facilitate both the sending of messages to other users (either at the same site or at other sites on the ARPA network) and the transfer of data and programs from site to site on that network. Also, the ability to "link" two or more terminals allows users to communicate easily and immediately.

Both the linking ability and the message facility have been found to be invaluable aids in inter-group communications and in such problems as interactive program debugging. When two terminals are linked, their output streams are merged, thus allowing each terminal to display everything typed at the other terminal. Since only the output stream is affected under these circumstances, it is still possible for each terminal to be used to provide input to separate programs, in addition to being used in a conversational mode.

Resource allocation. Some extensions to the

Research and development at SUMEX-AIM

The community of participants in SUMEX-AIM can be divided geographically into local (*i. e.*, Stanford-based) projects and remote projects. Below we give a brief description of the major representatives of each group of projects and of the SUMEX-AIM system, which is itself a significant development effort.

Communication with the remote projects is accomplished through one or both of the communications networks shown in the chart. In most cases, connection with SUMEX-AIM from these remote sites involves only a local telephone call to the nearest network "node." In this discussion we emphasize the medical implications of various research projects. In many instances, these potential applications serve as working examples of more-fundamental studies of AI research as well.

The Rutgers Project—Computers in Medicine. Professor Saul Amarel, a Rutgers University computer scientist, directs several research efforts designed to introduce advanced methods in computer science—particularly in artificial intelligence and interactive data base systems—into specific areas of biomedical research.

For example, a group of computer scientists led by Professor Casimir Kulikowski is developing computer-based consultation systems for diseases of the eye in collaboration with Dr. Aran Safir, an ophthalmologist from the Mount Sinai School of Medicine. An important development in this area is the establishment of a national network of collaborators for computer diagnosis and treatment of glaucoma.

The computer system, which includes an elaborate pathophysiologic model of the disease, is being tested through the SUMEX-AIM network at three eye centers: Mount Sinai Hospital and Medical Center, New York; Washington University, St. Louis; and The Johns Hopkins University, Baltimore.

Glaucoma, in one form or another, affects 2% of all people over 40 years of age. It is a disease in which increased pressure within the eye may lead to irreparable optic nerve damage and blindness. The computer-based program has great potential for assisting clinicians and researchers in understanding the disease, diagnosing it more accurately, and improving its treatment.

In another project, Professor Charles Schmidt, a social psychologist, is developing a theory of how people arrive at their interpretation of the social actions of others. The theory will be tested in situations such as the psychiatric interview and the legal trial. The computer system which currently represents the theory is called "Believer." It includes a large body of statements about people's motivations and actions. The SUMEX-AIM environment provides an excellent medium for collaboration between Dr. Schmidt's group and other researchers around the country in the development and testing of this computer-based theory.

The Rutgers project includes, in addition, several fundamental studies in artificial intelligence and system design. These provide much of the support needed for the development of complex systems such as the glaucoma consultation and the "Believer" programs.

DIALOG. The DIAgnostic LOGic project, under the direction of Drs. Harry Pople and Jack Myers at the University of Pittsburgh, is a large-scale, computerized medical diagnostic system utilizing the methods and structure of artificial intelligence. Unlike most computer diagnostic programs, which are oriented to differential diagnosis in a rather limited area, the DIALOG system deals with the general problem of diagnosis in internal medicine and currently accesses a medical data base encompassing approximately 50% of the major diseases in internal medicine.

MISL. The Medical Information Systems Laboratory at the University of Illinois at Chicago Circle has been established under the direction of Dr. Bruce McCormick, Information Engineering, in collaboration with Dr. Morton Goldberg, an ophthalmologist at the U of I medical school. The project explores inferential relationships between analytic data and the natural history of selected eye diseases, both in treated and untreated forms. SUMEX-AIM will be utilized to build a data base to be used as a test bed for the development of clinical decision support algorithms.

Computing Applied to Protein Crystallography. Members of the artificial intelligence project at Stanford also are collaborating with Professor Joseph Kraut and Dr. Stephan Freer, protein crystallographers at the University of California, San Diego. They are using the SUMEX-AIM facility as the central repository for programs, data, and other information of common interest.

The general objective of the project is to apply problem-solving techniques, which have emerged from artificial intelligence research, to the well-known "phase problem" of x-ray crystallography in order to determine the three-dimensional structures of proteins. It is intended to be of practical and theoretical value to computer science (particularly artificial intelligence research) and protein crystallography.

basic TENEX system have been made to facilitate resource management. Users are members of groups working on specific projects. It is among these projects that the facility is apportioned. Disk space and CPU cycles are now distributed among groups instead of among individual users.

For example, a user may exceed his individual disk allocation somewhat without any ill effect, so long as the total allocation of his group remains within the limits. Similarly, a Reserve Allocation Scheduler has been added to TENEX which tries to match CPU access to the schedule over a 90-sec time frame. Thus a particular group cannot dominate the machine if a lot of its members are logged in at one time.

It is typical for usage of a facility to peak through the middle hours of the day. Indeed, one of the advantages of having users from around the country is the spreading of the load caused by the difference in time zones. To further optimize usage, SUMEX-AIM publishes a weekly plot of diurnal loading. This plot shows the total number of jobs on the system as well as the number of LISP jobs, since these jobs seem to make the biggest demands of system resources. The result has been an increased awareness by users of system loading and a noticeable increase in the number of users at all hours of the night and early morning.

Protection and system security. Protection for a computer system covers a range of ideas. It means the ability to maintain secrecy —for example, to guarantee the privacy of patient records. It also guarantees integrity by assuring that programs and data are not modified by an unauthorized party. Since it is primarily an academic research facility, the data in which are unlikely to attract heroic efforts of penetration, SUMEX-AIM has not had to resolve the very difficult problems often raised by commercial and military use.

Tenex provides the usual keyword protection at login time and a measure of file protection. Owners of a file may assign a protection number which specifies some combination of *read, write, execute,* or *append* access to a file for owners, members of a group, or other users. This level of protection is basically enough to prevent accidents and most mischief.

TENEX passwords are, however, stored within the monitor program. Hence, additional facilities allow users to encrypt and decrypt any of their files with specific password protection for each file, invoking algorithms that do not store an explicit representation of the key. Users who would require even greater data security perhaps have an even more formidable problem in finding hardware protection against wiretapping and similar physical intrusions.

In an effort to improve the human engineering of programs for public use, the capability of recording a session has been built into several of the programs. Studied by the program designers to pinpoint confusing aspects of programs, these recordings serve to improve program design. Some of these programs now request permission to record a session before doing so.

File backup. In order to assure the user maximum protection against loss of valuable work, SUMEX operates a multi-level file backup system. In addition to a routine file backup system there are facilities to enable the user to selectively archive disk files.

A simple command to the TENEX executive requests the operator to copy specified files to magnetic tape. Each such file is copied to two magnetic tapes within 24 hours of issuing the archive command. File retrieval is affected by a similar process.

The user also has the alternative option of being able to lodge files in a special backup directory. Files are held in this directory until the next exclusive file dump (see below) at which time they are deleted. In this way the user can remove files from his directory at his own choosing, knowing they will be archived by the exclusive dump.

On a system level, an effort is made to maintain file backups such that the maximum possible loss, in the event of a crash fatal to the file system, would amount to no more than one day's work. Once each day all files that have been read or written within the last 48 hours are dumped onto magnetic tape. Files that exist for 48 hours are thus held on two separate tapes. The rotation period for files dumped in this way is 60 days.

Once each week a full file dump is made to separate disk storage. Each such dump is kept for two weeks at which time it is replaced by a new file dump. Each month there is a full system dump from disk to magnetic tape. Files can be recovered from the system backup by sending a message to the operator specifying the file name(s) and when the file was last read or written (if such information is available).

Conflicting requirements of research and production programs. One of the concepts behind the creation of a shared resource is elimination of the problems which arise when large, complex computer programs are exported. Since, in theory, exportability is no longer a problem, there is greater latitude in choice of a language in which program development can take place. In the case of some of the DENDRAL programs, it was thought that program development should take place in INTERLISP, a language that lends itself well to the artificial intelligence nature of these programs, but does not lead to particularly efficient run-time code.

In order to ascertain the usefulness of these programs and to determine what areas remain in need of work, chemist collaborators are being sought. As these users increase in number and begin to use the programs more frequently, it is almost certain that the inherent slowness of the predominately LISP code will affect the whole system as well as handicap the efficient use of the DENDRAL programs.

Additionally, some of the chemist-users who are finding the programs most useful and who are most enthusiastic about their

Stanford projects

MYCIN—Computer Based Consultation in Clinical Therapeutics. Dr. Stanley Cohen, Professor and Head of the Division of Clinical Pharmacology at Stanford, directs this research in collaboration with Dr. Stanton Axline and with computer scientists interested in artificial intelligence and medical computing.

An evolving computer program developed to assist physician nonspecialists in the selection of therapy for patients with bacterial infections, MYCIN attempts to model the decision processes of medical experts. It consists of three closely integrated components:

- the Consultation system asks questions, makes conclusions and gives advice;
- the Explanation system answers questions from the user to justify the program's advice and explain its methods;
- the Rule-Acquisition system permits the user to teach the system new decision rules or to alter preexisting rules judged to be inadequate or incorrect.

Goals for further development of the system include expansion of the consultation program to deal with infections other than bacteremias and implementation and evaluation of the system in the clinical setting at Stanford University Hospital.

DENDRAL. The DENDRAL project at Stanford under the direction of Professors Lederberg; Edward Feigenbaum, Computer Science; and Carl Djerassi, Chemistry; is aimed at assisting the biochemist in interpreting molecular structures from mass spectral (3-5) and other chemical information. In cases where the characteristic spectrum of a compound is not cataloged in a library, the DENDRAL programs carry out the rather laborious processes a chemist must go through to interpret the spectrum from "first principles."

By symbolically generating "reasonable" candidate structures from hints within the spectrum and a knowledge of organic chemistry and mass spectrometry, the program infers the unknown structure to be the one which best explains the observed spectrum. There is no direct algorithmic path available to determine such a molecular structure from the spectral data—only the inferential process of hypothesis generation and testing within the domain of reasonable solutions defined by a knowledge of organic and physical chemistry.

This process, as implemented in the computer, is a simplified example of the cycle of inductive hypothesis—deductive verification that often is taught as a model of the scientific method. (Whether this is a faithful description of contemporary science is arguable, and how it may be implemented in the human brain is unknown. Regardless, these are useful leads rather than absolute preconditions for the pragmatic improvement of mechanized intelligence for more efficient problem-solving.) The elaboration of these approaches with existing hardware and software technologies is the most promising approach to enhancing computer application to the vaguely-structured problems that dominate our task domains.

SUMEX-AIM system development

Current research activities at SUMEX-AIM are developing along several lines. On a system development level there are ongoing projects designed to make the system more user oriented. Currently, the system can be expected to provide help to the user who is confused about what is expected in response to a certain prompt from the computer. A "?" typed into the computer terminal by the user, will, in most cases, provide a list of possible responses from which to choose. Also available, in response to typing "HELP" into the monitor, is a general help description with a list of files likely to be of interest to a new user.

Another area of system development currently being explored at SUMEX-AIM is that of creating a comprehensive "bulletin board" facility where users can file "bulletins," that is, messages of interest to the SUMEX-AIM community. The facility also will alert users to new bulletins which are likely to be of interest to them, as determined by individual user-interest profiles.

An additional aspect of our research is our commitment to share developments among a wider community. For example, several of the DENDRAL programs are advanced enough to be useful to chemists engaged in related work in mass spectrometry and structure elucidation in general. These programs are written primarily in the programing language INTERLISP (6), and thus are not easily exportable.

SUMEX-AIM provides a mechanism for allowing others access to the programs without the requirement for any special programing or computer system expertise. The availability of the SUMEX facility over nationwide networks allows remote users to access the programs, in many instances via a local telephone call.

These DENDRAL programs have only recently been released for outside use. Some announcement of their availability has been made, and other announcements will occur in the near future, through talks, publications in press, demonstrations, and informal potential use, are persons working in industry. This interest from industry could be interpreted as an indication of the "real-world" usefulness of the programs. Special arrangements are under study to solve the problem of offering access to SUMEX, on a fee basis, on the part of private industry.

Community mindedness. Those involved in computer science research at SUMEX face a general problem which is absent or greatly lessened at non-network sites: the problem of community mindedness. The network provides a large and varied set of other researchers and users who have an interest in their work. Although the network-TENEX combination provides new forms of communication with these remote parties, the traditional means of fully describing the use and structure of a complex program, a detailed person-to-person discussion, is not convenient.

Comprehensive documentation gains importance in such a situation, and within the DENDRAL project considerable time has been needed in the development of program descriptions which are adequate for a diverse audience. Also, in both DENDRAL and MYCIN, effort has been and is being directed toward "human engineering" in program design: to provide the user with commands which assist him in using the programs, in understanding the logic by which the programs reach certain decisions, and in communicating questions or comments on the programs' operation to those responsible for development. Such "housekeeping" tasks are often neglected, yet are quite important in smoothing interaction with the community.

Choice of programing language. High level programing languages which are designed for ease of program development are frequently poor as production-level languages. This is because developmental languages free the researcher from a raft of programing details, thus allowing him to concentrate upon the central logical issues of the problem, but the automatic handling of these details is seldom optimal.

Also, because such languages tend to be specialized for certain computers and operating systems, the exportation of programs can be a serious problem. One solution to these problems is the recoding of research-level programs into more-efficient language when fast and exportable versions are needed.

Networking greatly eases the problem of exportability, but can also aggravate the problem of efficiency. As mentioned in the previous section, the DENDRAL programs, which are undergoing constant development, found a substantial number of production-level users. Because of the inefficiencies of INTERLISP (a 50- to 100-fold improvement in running time is not uncommon when an INTERLISP program is translated into FORTRAN), the use of INTERLISP adversely impacted the entire system.

Because the DENDRAL programs are quite large and complex, their translation into other languages is impractically tedious. A partial solution to this problem is provided by the TENEX operating system, which allows some interface between programs written in different languages.

With such intercommunication, time-consuming segments of an INTERLISP program which are not undergoing active development can be reprogramed in another more efficient language. The developmental parts of the program are left in INTERLISP, where modifications can easily be made and tested.

The CONGEN program uses three languages: INTERLISP, FORTRAN, and SAIL. The SAIL segment was added when a new feature, whose implementation was fairly straightforward, was included in CONGEN. Since then, the SAIL portion gradually has been taking over some of the more time-consuming tasks. This method allows a balance in the tradeoff between ease of program development and efficiency of the final program.

Accumulation of expert knowledge in knowledge-based programs. Just as statistics-based programs need to worry about accumulation of large data bases, knowledge-based programs need to worry about the accumulation of large amounts of expertise. The performance of these programs is tied directly to the amount of knowledge they have about the task domain—in a phrase, knowledge is power. Therefore, one of the goals of artificial intelligence research is to build systems that not only perform as well as an expert but that also can accumulate knowledge from several experts.

The networking aspects of SUMEX-AIM allow a wide range of expert input from dispersed locations. This is shown in the Rutgers project research on glaucoma, linking eye centers in New York, St. Louis, and Baltimore. Regional differences in expertise are more characteristic of research in clincial medicine than the basic medical sciences.

Simple accretion of knowledge is possible only when the "facts," or inference rules, that are being added to the program are entirely separate from one another. It is unreasonable to expect a body of knowledge to be so well organized that the facts or rules do not overlap. (If it were so well organized, it is unlikely that an artificial intelligence program would be the best encoding of the problem solver.)

One way of dealing with the overlap is to examine the new rules on an individual basis, as they are being added to the system, in order to remove the overlap. This was the strategy for developing the early DENDRAL programs. However, it is very inefficient and becomes increasingly more difficult as the body of knowledge grows.

The problem of removing conflicts, or potential conflicts, from overlapping rules becomes more acute when more than one expert adds new rules to the knowledge base. Of course, the advantages of allowing several experts to "teach" the system are enormous—not only is the program's breadth of knowledge potentially greater than that of a single expert, but the rules are more apt to be refined when looked at by several experts. On

discussions. Although most of our experience has been with local users, they have been good models of remote users in that their previous exposure to the actual programs and computer systems is minimal.

Their participation has been extremely useful in helping us to smooth out clumsy interactions with programs and to locate and fix program bugs. Such polishing is important for programs which may be used by persons from widely differing backgrounds with respect to computers, networks, and time sharing systems.

We are in the processes of building a community of remote users. We actively encourage such use for two reasons:

■ we feel the programs are capable of assisting others in solving certain molecular structure problems, and

■ such experience with outside users will be a tremendous assistance in increasing the power of our programs as the programs are forced to confront new real-world problems.

CONGEN (7-13) is one of the programs in the DENDRAL project that is providing this networking experience with outside users.

Structure problems are usually not solved with mass spectrometry alone. Even when sample size is too limited for obtaining other spectroscopic data, knowledge of chemical isolation and results of derivative formation procedures frequently act as powerful constraints on structural possibilities. Larger amounts of sample permit determination of other spectroscopic data.

Taken together, this information allows determination of structural features (substructures) of the molecule and constraints on the plausibility of ways in which the substructures may be assembled. The CONGEN program is capable of providing assistance in solution of such problems.

In an interactive session with the program, a user supplies structural information determined by his own analysis of the data (perhaps with the help of the above programs). In addition, the user supplies whatever other constraints are available concerning desired and undesired structural features, ring sizes and so forth.

The program builds structures in a series of steps, during which a user can interact further with the procedure, for example, to add new constraints. Although very much a developing program, its ability to accept user-inferred constraints from many data sources makes CONGEN a general tool for structure elucidation which we are making available via SUMEX-AIM in its current form. ■

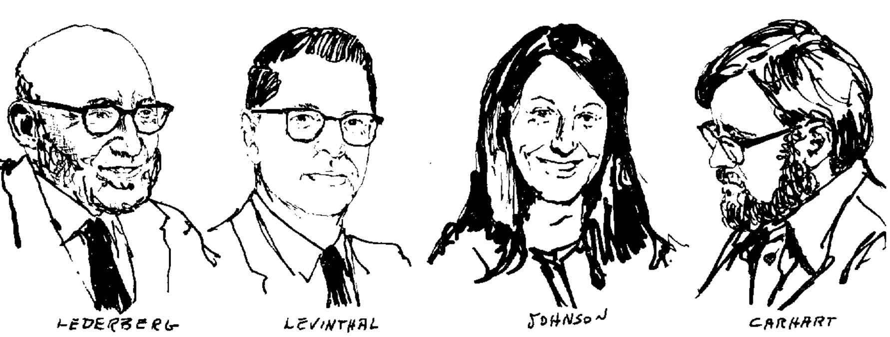
THE AUTHORS

Elliott C. Levinthal, director of the instrumentation research laboratory, also is adjunct professor, Dept. of genetics, Stanford Univ. School of Medicine. He holds a PhD from Stanford. Special fields are design of microwave tubes, nuclear resonance, medical electronics, and the application of computers to image processing and medical information systems.

Raymond E. Carhart, PhD in chemistry, Calif. Inst. of Tech., is research associate in computer science, dept. of genetics. He is working on the CONGEN program (see text) and computerized analysis of ^{13}C NMR and mass spectra.

Suzanne M. Johnson, BS chem, Univ. of Ariz., is applications programer with SUMEX, providing a user interface to DENDRAL programs. She has published extensively in x-ray crystallography.

Joshua Lederberg, chairman, dept. of genetics, Stanford Univ., is principle investigator SUMEX-AIM. Among his many accomplishments, he is widely respected for work on the organization of genetic material in bacteria (Nobel Prize in 1958) and the chemistry and evolution of unicellular organisms and of man. His PhD (Yale) has been followed by honorary degrees from numerous respected universities.



the other hand, one can expect not only a greater volume of new rules but a higher percentage of conflicts when several experts are adding rules.

Having a computer program that can accumulate knowledge presupposes having an organization of the program and its knowledge base that allows accumulation. If the knowledge is built into the program as sequences of low-level program statements—as often happens—then changing the program becomes impossible. Thus current artificial intelligence research stresses the importance of separating problem-solving knowledge from the control structure of the program that uses that knowledge.

Another problem, at a political rather than a programing level, becomes apparent with one accumulation process: how does the program distinguish an expert from a novice? This problem is made more important because of networking. The users are apt to be less well-known to the program designer.

In the MYCIN program we have circumvented the problem by having the program ask the current user for a keyword that would identify him as an expert. It is then a bureaucratic decision as to which users are given that keyword. There is nothing subtle in this solution, and one can imagine far better schemes for accomplishing the same thing. The point here is that not every user should have the privilege of changing rules that experts have added to the system, and that some safeguards must be implemented.

The "security" of a local facility. Networking is still a relatively new concept to many people, and there is a resistance to departing from the "traditional" modes of computing. There is a sense of security in having a local computing facility with knowledgeable consultants within walking distance, and in having "hard" forms of input (*e.g.*, boxes of computer cards) and output (*e.g.*, voluminous listings).

These props are difficult to simulate over a network connection—in most cases a user's interaction with the remote site takes place exclusively through a computer terminal—yet the quality of service can match or exceed that of a local facility. Programs and large data sets can be entered and stored on secondary storage as can large output files. All types of program and data editing can be done with interactive editing programs. Programs can be written in an interactive mode so that small amounts of control information can be input and key results output in "real time" over the terminal. And as noted in a previous section, consultation can be significantly more productive, providing that the remote operating system supports the appropriate types of communication possibilities.

There can, of course, be no denying that there are problems in learning to use a distant computer system, be it for program development or for the use of certain programs. Only the individuals involved can say whether or not overcoming these problems, to gain access to the special resources which are available, is worth the effort. Fortunately, there will always be those persons who have a pressing problem in need of solution and who are willing to try a new approach— regardless of whether or not they have had prior network experience. ■

Notes and references

The SUMEX-AIM facility is supported by the National Institutes of Health (grant number RR 00785), as is the DENDRAL project (grant number RR 00612-05A1).

1. Professor Joshua Lederberg, Principle Investigator.
2. Carhart, R.E., *et al*, "Networking and a Collaborative Research Community: A Case Study using the DENDRAL PROGRAMS," in "Computing, Networking, and Chemistry." ACS Symposium Series, No. 19, 1975.
3. Lederberg, J., *et al*, J. ACS (1969), 91, 2973.
4. Duffield, A.M., *et al*, J. ACS (1969), 91, 2977.
5. Buchanan, B. G., Duffield, A. M., and Robertson, A. V., "Mass Spectrometry: Techniques and Applications." John Wiley & Sons, New York, 1971.
6. Teitleman, W., "INTERLISP Reference Manual," Xerox Corp., Palo Alto, Cal., 1974.
7. Smith D. H., and Carhart, R. E., Abstracts, 169th Meeting ACS, April 6-11, 1975.
8. Carhart, R. E., *et al*, J. ACS, submitted for publication.
9. Masinter, L. M., *et al*, J. ACS (1974), 96, 7702.
10. Masinter, L. M., *et al*, J. ACS (1974) 96, 7714.
11. Brown, H., SIAM Journal of Computing, submitted for publication.
12. Carhart, R. E., Smith, D. H., and Brown, H. J., Chem. Inf. Comp. Sci., in press (May, 1975).
13. Smith, D. H., Anal Chem., in press (May, 1975).
14. Bobrow, D. G., Burchfiel, J. D., and Tomlinson, R. S., Commun. ACM (1972), 15(3), 135.
15. VanLehn, K. S., "SAIL User Manual," Stanford Artificial Intelligence Laboratory, Stanford, Cal., 1973.